MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, 2019/2020

DPS5211 –INTRODUCTION TO PROBABILITY AND STATISTICS

(For Diploma students only)

14 MARCH 2020
9.00 a.m - 11.00 a.m
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This question paper consists of 8 pages (4 pages with 4 questions and 4 pages for appendix)
2. Answer **ALL** questions. All necessary working steps must be shown.
3. Write all your answers in the answer booklet provided.
4. Statistical Tables will be provided.

QUESTION 1 [10 MARKS]
[FOR PART (a) iv., (c) and (d), CORRECT YOUR ANSWERS TO FOUR DECIMAL PLACES.]

The time taken in seconds by a group of students in a class to complete a quiz are listed as follows.

31, 34, 41, 33, 29, 31, 39, 35, 30, 40.

(a) Use the given list as ungrouped data to find

i.	the mean time.	[1 mark]
ii.	the median time.	[1 mark]
iii.	the mode.	[0.5 marks]
iv.	standard deviation of the sample.	[2.5 marks]

(b) Calculate the interquartile range. [1 mark]

(c) Find the probability that a student chosen at random from the group took at least 35 seconds to complete the exercise. [3 marks]

(d) Find the probability that a student chosen at random from the group took exactly 29 seconds to complete the exercise. [1 mark]

Continued...

QUESTION 2 [10 MARKS]
[CORRECT YOUR ANSWERS TO FOUR DECIMAL PLACES.]

According to a study, the likelihood for the number of cars in Kuala Lumpur household is given in the table form as below:

Number of cars (x)	0	1	2	3	4
Probability $P(X = x)$	0.03	0.13	0.70	0.10	0.04

(a) Find $P(X \leq 2)$. [1 mark]

(b) For the above probability distribution, find
i. the mean. [1 mark]
ii. the standard deviation of this distribution. [2 marks]

(c) A small digital storage manufacturer makes novelty USB drives which are given away at conferences. The probability that a drive is defective is 0.03. The factory tests a random sample of 200 drives.

i. Find the probability that there are exactly ten defective drives in the sample. [1 mark]
ii. Find the probability that there is at least one defective drive in the sample. [2 marks]

(d) The probability of a telesales representative making a sale on a customer call on average is 0.15.

i. Find the probability that no sales are made in 10 calls. [1 mark]
ii. Find the probability that at most 2 sales are made in 10 calls. [2 marks]

Continued...

QUESTION 3 [10 MARKS]
[CORRECT YOUR ANSWERS TO FOUR DECIMAL PLACES.]

(a) A survey of 500 airline passengers found that 338 were satisfied with the service they received from the flight attendants.

i. Find the point estimate of the population proportion. [1 mark]
ii: Calculate and interpret a 95% confidence interval for the proportion of passengers who are satisfied with the service from flight attendants. [2 marks]

(b) Pepsi's dispensing machine is designed to fill bottles with exactly 2 liters of their product. To test if the machine is performing according to specification, a sample of 100 "2-liter" bottles are collected. The average quantity contained in the sample bottles is 1.985 liters. The population standard deviation of the fill is known to be 0.05.

Test whether the machine is in control, at 5% level of significance. [3.5 marks]

(c) The training department of a company wishes to determine if there is any difference in the performance between the workers who have completed a training program and those who have not completed the program. A sample of 100 trained workers have given better performance, with an average key performance index (KPI) of 74.3 and a sample standard deviation of 16. A sample of 100 workers who have not been trained, did not give good performance and the average KPI is 69.7 with a standard deviation of 18.

i. State the null and the alternative hypothesis. [1 mark]
ii. By using the hypothesis testing of two population mean, calculate the test statistic. [1 mark]
iii. Test whether there is evidence of difference in performance between the two groups at 5% significance level. [1.5 marks]

Continued...

QUESTION 4 [10 MARKS]
[FOR (b) and (c), CORRECT YOUR ANSWERS TO FOUR DECIMAL PLACES.]

The table below shows the scores for 12 students on two Mathematic exam papers. For the first paper, calculators were allowed and for the second paper, they were not.

Paper 1 (x)	74	73	65	75	68	72	69	71	83	68	68	73
Paper 2 (y)	75	83	69	77	71	77	68	76	84	69	71	75

(a) Calculate $\sum x$ and $\sum y^2$. [2 marks]

(b) Find the value of SS_{yy}. [2 marks]

(c) Referring to part (a) and part (b) and given, $SS_{xx} = 240.9167$ and $SS_{xy} = 228.9167$, find the simple linear regression equation that fits the data obtained. [3.5 marks]

(d) Find the marks in paper 2, when the mark in paper 1 is 85. [1 mark]

(e) Compute and interpret the sample correlation coefficient, r. [1.5 marks]

Continued...

APPENDIX – KEY FORMULA

- **Mean for ungrouped data**

$$\bar{x} = \frac{\sum x_i}{n} \quad \text{or} \quad \mu = \frac{\sum x_i}{N}$$

- **Mean for grouped data**

$$\bar{x} = \frac{\sum x_i f_i}{n} \quad \text{or} \quad \mu = \frac{\sum x_i f_i}{N}$$

Where x is the midpoint and f is the frequency of the class

- **Variance for ungrouped data**

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \text{or} \quad s^2 = \frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}$$

- **Variance for grouped data**

$$s^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{n-1} \quad \text{or} \quad s^2 = \frac{\sum x_i^2 f_i - \frac{\left(\sum x_i f_i\right)^2}{n}}{n-1}$$

Where x is the midpoint and f is the frequency of the class

DISCRETE RANDOM VARIABLE, BINOMIAL DISTRIBUTION, AND POISSON DISTRIBUTION

- Mean of a discrete random variable x : $\mu = \sum xP(X = x)$
- Standard deviation of a discrete random variable x : $\sigma = \sqrt{\sum x^2 P(X = x) - \mu^2}$
- Binomial probability formula : $P(X = x) = \binom{n}{x} p^x q^{n-x}$
- Poisson probability formula : $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

- Key formula for Binomial Probability Distribution:
  - If $p \leq 0.5$
    i. $P(X = r) = B(r) - B(r-1)$
    ii. $P(X \geq r) = 1 - B(r-1)$
    iii. $P(X \leq r) = B(r)$
    iv. $P(X > r) = 1 - B(r)$
    v. $P(X < r) = B(r-1)$
    vi. $P(a \leq X \leq b) = B(b) - B(a-1)$

  - If $p > 0.5$
    i. $P(X = r) = B(n-r) - B(n-r-1)$
    ii. $P(X \geq r) = B(n-r)$
    iii. $P(X \leq r) = 1 - B(n-r-1)$
    iv. $P(a \leq X \leq b) = B(n-a) - B(n-b-1)$

- Key formula for Poisson Probability Distribution:
  - If $\mu = \lambda$
    i. $P(X = r) = Poi(r) - Poi(r-1)$
    ii. $P(X \geq r) = 1 - Poi(r-1)$
    iii. $P(X \leq r) = Poi(r)$
    iv. $P(X > r) = 1 - Poi(r)$
    v. $P(X < r) = Poi(r-1)$
    vi. $P(a \leq X \leq b) = Poi(b) - Poi(a-1)$

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value : $Z = \frac{x - \mu}{\sigma}$

ESTIMATION

- The $(1-\alpha)100\%$ confidence interval for μ is

$\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}$ where $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

$\bar{x} \pm t_{\alpha/2,n-1}s_{\bar{x}}$ where $s_{\bar{x}} = \dfrac{s}{\sqrt{n}}$

- The $(1-\alpha)100\%$ confidence interval for p is

$\hat{p} \pm z_{\alpha/2}s_{\hat{p}}$ where $s_{\hat{p}} = \sqrt{\dfrac{\hat{p}\hat{q}}{n}}$

- **Sample size**

$$n = \left[\frac{\left(Z_{\alpha/2}\right)\sigma}{E}\right]^2 \qquad n = \frac{\left(Z_{\alpha/2}\right)^2 pq}{E^2}$$

HYPOTHESIS TESTING

- **ONE population**

Mean:

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \quad \textbf{or} \quad t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

Proportion:

$$Z = \frac{\hat{p}-p}{\sqrt{\dfrac{pq}{n}}}$$

- **TWO population**

Mean:

$$Z = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^{\,2}}{n_1}+\dfrac{\sigma_2^{\,2}}{n_2}}}$$

Proportion:

$$Z = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\hat{p}_p(1-\hat{p}_p)}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \qquad \text{where} \qquad \hat{p}_p = \frac{x_1+x_2}{n_1+n_2}$$

REGRESSION

- For the **least square regression equation,** $\hat{y} = b_0 + b_1 x$

  where $b_1 = \dfrac{SS_{xy}}{SS_{xx}}$ and $b_0 = \dfrac{1}{n}\left[\sum y - b_1 \sum x\right]$

- **Correlation coefficient,** r

  $$r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}}$$

  where

  $$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n}, \qquad SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} \text{ and}$$

  $$SS_{xy} = \sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n}$$

- **Regression *t*-Test**

  $$t = \frac{b_1 - \beta_1}{s_b} \qquad \text{where } s_b = \frac{s_e}{\sqrt{SS_{xx}}} \qquad \text{and } s_e = \sqrt{\frac{SS_{yy} - b_1 SS_{xy}}{n-2}}$$

- **Correlation *t*-Test**

  $$t = \frac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$$

End of Paper.